The Tenth Graham Clark Lecture

THE SCIENCE OF NEAR SPACE

By Sir Edward Appleton, G.B.E., K.C.B., M.A., D.Sc., F.R.S. (*Honorary Member*)*

TO BE INVITED to give this lecture is an honour I warmly appreciate. I had the privilege of knowing Graham Clark whom we remember in gratitude on this annual occasion. While he was, of course, responsible for ensuring that the business of our oldest engineering institution was conducted with efficiency and despatch, he was at the same time intensely interested in men and ideas. It was, indeed, this rich humanity which extended the ambit of his friendships far beyond the immediate circle of the institution which he loved, and served, with such devotion and distinction.

Reflections of this kind are responsible for the choice of the theme of my tribute tonight. For I have chosen to talk about a topical subject, one which could so readily have been a matter of conversational discussion between Graham Clark and myself; for his scientific curiosity was inexhaustible.

That the relations between the sun and the earth are indeed a topical subject is illustrated by the fact that, over the period of this year and next, during what have been picturesquely called the two International Years of the Quiet Sun, geophysicists the world over are to conduct an internationally agreed programme of scientific observations, designed to elucidate solar and terrestrial relationships. Putting things in another way, I could also say that my subject is a limited aspect of space research, because, as we shall see, the terrestrial effects of emitted solar particles are markedly dependent on the conditions that they encounter, in space, as they travel from the sun to the earth.

There are two further matters that I think I should mention by way of preamble. The subject of near space, of the relation between the sun and the earth, has been advanced by scientific workers of many nations. Undoubtedly, both Russia and the United States have been responsible for some of the more spectacular aspects of space research, so far as the launching of earth satellites and space-probes is concerned. But workers in other countries have also been responsible for new ideas, and even for the design and construction of new equipment which has been launched aloft by way of American facilities.

The second feature of these recent developments I want to stress is the way that science and engineering have worked in partnership. Not only have we witnessed enormous mechanical engineering developments in the launching of huge vehicles into space and their subsequent orientation and control in desired orbits; there have also been revolutionary developments in electronic circuit design with miniaturized components. In the latter connection it is scarcely necessary for me to remind an engineering audience that radio communication must be used for the conveyance of instructions to a space-craft as well as for the transmission of the information it gathers to receiving stations on the ground.

Now, although undoubtedly there have been remarkable advances in our knowledge of the conditions in the outer fringe of our own atmosphere and beyond since measurements became possible *in situ*, it would be a distortion of the history of space research to underrate the accumulation of knowledge achieved by way of earth-bound instruments, long before Sputnik I was launched in 1957. However, when earth-bound instruments were the only means of observation we were restricted to making inferences only concerning the physical conditions at places away from the earth's surface. When measurements have more recently been made on the spot, so to speak, these inferences have mostly been confirmed in a general way. But, usually, quite unexpected discoveries have also accrued from measurements made *in situ*. To two recent discoveries of this kind I now turn.

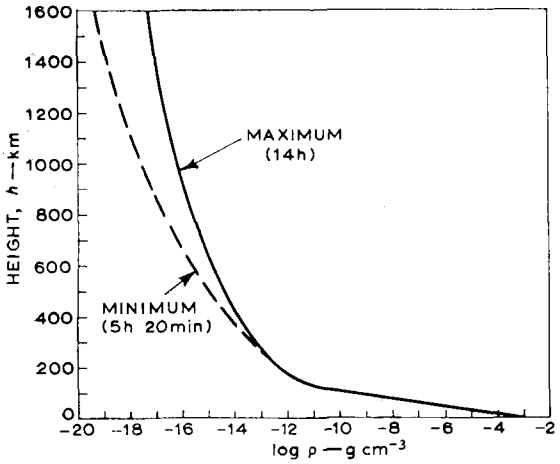
The first example I want us to consider is concerned with the answers to the question: How far out from the earth's surface does our terrestrial atmosphere extend? We know how rapidly the atmospheric density decreases with increase of height in that part of the atmosphere

This lecture was delivered at a Joint Meeting of the Institutions of Civil, of Mechanical, and of Electrical Engineers, held at the Institution of Civil Engineers, London, on 24th March 1964.

* *Principal and Vice-Chancellor of the University of Edinburgh.*

which is near the ground and so is accessible. This leads us to conclude that, at a height of 100 km above ground, we would expect the air density to be just about one millionth of its value at ground level. But suppose we consider still greater heights, assuming that the same factor of dilution for each 100 km is appropriate. This would lead us to conclude that at a height of 300 km the air density would be 10^{-18}—a millionth of a millionth of a millionth—of its earth-surface value. This is another way of saying that we should expect the population density of air molecules at 300 km to be only 10 per cubic centimetre, a very small value indeed. However, geophysical researches carried out long before the advent of earth satellites had already shown that such a low molecular population density as 10 molecules per cubic centimetre could not possibly obtain at a level of 300 km above ground. For one thing, we know, from the radio sounding of the upper atmosphere, that there must be, during daytime, at least a million electrons per cubic centimetre at a level of 300 km which is approximately the level of the maximum of the ionospheric F layer. This figure is in flagrant disagreement with the estimate I have just mentioned, of 10 molecules per cubic centimetre, for the molecular density. From this we infer that there are some physical processes operative which cause the atmosphere to decay more and more slowly as we go to greater and greater heights, the effect of which is to make the molecular population density still quite appreciable at very great heights (Fig. 1).

Now this picture of the extension of the earth's atmosphere to very great heights has been beautifully confirmed in recent years by observations made on the orbital timing of earth satellites. Any such satellite which approaches to within about 1000 km of the earth's surface is notably affected by air drag, which reduces its energy on each revolution and thereby makes its orbit contract. From such observations on orbital movement it is possible to deduce the density of the air in the small section of the satellite's orbit near perigee. Now from a great wealth of measurements of this kind we may be quite certain that the atmospheric density is still appreciable at very great heights, the value at 300 km being, not 10 molecules per cubic centimetre, but as much as 10^9 per cubic centimetre, which is one hundred million times as much. In fact these new observations confirm beautifully our old ideas about the extension of the atmosphere to great heights. Moreover, it is now quite possible to explain the presence of an electron density of a million electrons per cubic centimetre at the maximum of the F2 layer of the ionosphere if the molecular air density there is one thousand times the same numerical value. However, the orbital timing experiments have not merely confirmed old inferences; they have led to new discoveries. It has been shown that the air density at great heights, far from being constant, is subject to regular diurnal and annual changes, Fig. 2. But, perhaps even more unexpected, it has been



Note the extension of the atmosphere to great heights. The air above about 100 km is hotter, and of lower molecular weight, than at lower levels. Its rate of decay with height is therefore less. The two curves relate to different local times.

Fig. 1. Relation between air density and height above ground in the atmosphere (after Dieminger)

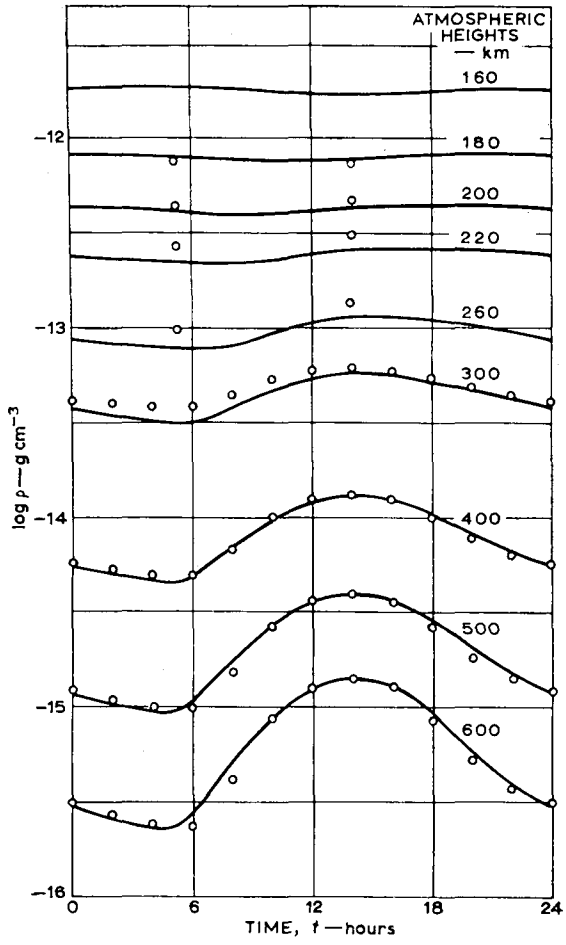


Fig. 2. Diurnal variations of air density at different atmospheric heights, as deduced from satellite orbital timing studies (after Harris and Priester)

found that the air density at great heights waxes and wanes with a secular variation in sympathy with the sunspot cycle.

My second example of the illuminating relevancy of new experiments to old problems relates to the occupancy of interplanetary space. For many years now we have inferred that the space between the earth and the sun is not always empty, not always totally devoid of matter. By analysing and interpreting observations made with earth-bound instruments a series of solar and terrestrial relationships have been established from which we have inferred that a primary cause of magnetic storms on the earth is the irruption, into the higher terrestrial atmosphere, of streams of charged corpuscles emitted by the sun. The additional fact that such corpuscles can cause auroral displays has led to the conclusion that they must be electrically charged. But, if this were so, the fact that they can travel from the sun to the earth along a localized track, without spreading, has indicated that the particles must consist of equal numbers of positively and negatively charged constituents. In other words it must be what we nowadays call a plasma.

Now the first successful attempt to observe the interplanetary plasma *in situ* was made with positive-ion traps carried on one of the Russian Lunik space probes. A flux of positive ions of about 2×10^8 particles per square centimetre per second was measured beyond a distance of 39 earth radii. However, the strongest evidence for the existence of a continuous flux of solar plasma, or solar wind in space, was obtained by means of a positive-ion spectrometer carried by the American space-probe Mariner 2. In this case the observations persisted over a period of a few months, so that it was possible to compare them with indications of magnetic storminess as measured on the earth over the same period. A most significant correlation was found in this way. The plasma velocity, which ranged from 400 kilometres per second to 700 kilometres per second over the period, was found to be intimately associated with the daily figures of magnetic storminess. When the plasma velocity was high the magnetic storminess was high, and *vice versa*, Fig. 3.

Now I think we can fairly claim that the first systematic measurements on conditions approaching near space began with the regular sounding of the ionosphere, by radio probing, in the nineteen-thirties. The ionosphere, as is now well known, is a shell of ionized air—a plasma —which surrounds the earth like orange peel round an orange. It is a realm of the universe which can be said to have been both invented and discovered. It was invented as an essential ingredient in theories formulated to account for the remarkable effectiveness of round-the-world communication by radio. One of its inventors in this connection was a Faraday Medallist of the Institution of Electrical Engineers, that great genius Oliver Heaviside. The subsequent discovery of the ionosphere took place in 1924 when it was located and probed by a technique we should now call frequency-modulation radar, a technique which is now finding an important application in the aircraft altimeter used in modern blind-landing systems. Since 1924, then, the ionosphere has been the subject of international exploration, as a result of which we have acquired a detailed picture of its world morphology, its diurnal and seasonal variations and also its variation in the course of the 11-year sunspot cycle. Such systematic observations, as I have already indicated earlier, have informed us about the ionospheric electron population density up to a height of about 300 km where the peak density of the ionosphere is found to occur. But what about electron population densities at levels well above 300 km? Does the ionosphere, like the atmosphere itself, extend to enormous heights until it blends into the plasma occupancy of space?

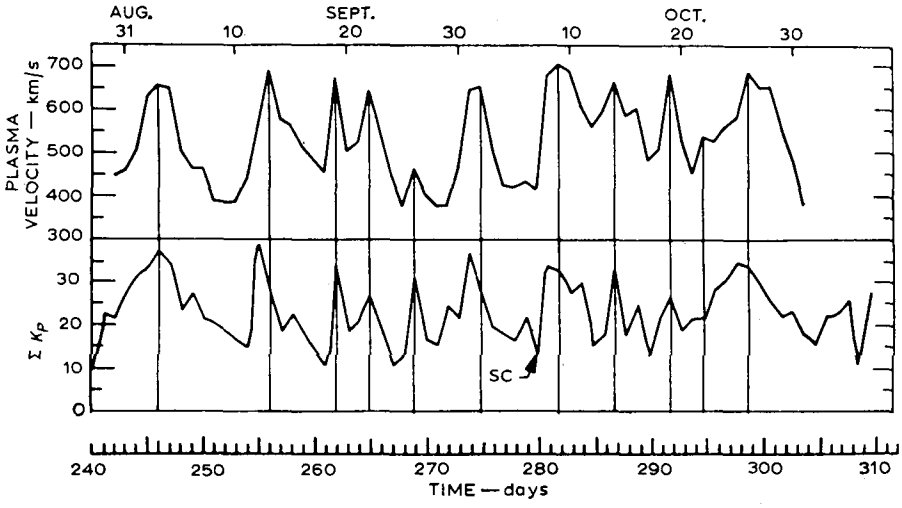
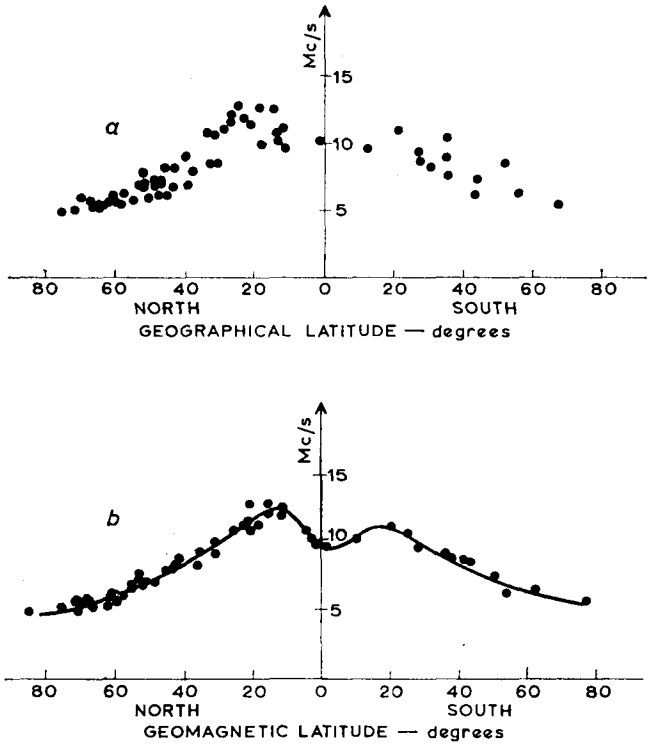


Fig. 3. Close correlation between solar wind velocity, as measured by an ion-counter in space, and magnetic storminess, as measured on the earth (after Snyder, Neugebauer, and Rao)

However, before I come to deal with the answers to these important questions, it is necessary for me to mention a characteristic of the ionosphere, disclosed by the normal radio probing method, which has been the subject of much attention in recent years. During the 1939–45 war I was much concerned with the subject of ionospheric forecasting. This was an attempt to forecast, for the benefit of radio engineers concerned with long-distance short-wave circuits, the optimum frequencies for use in different parts of the world over different ranges of communication. Usually such forecasts were made three months in advance; and, for this purpose, we had to forecast the peak electron density of the ionosphere all over the world at local different times. It was in the course of studying world maps of this peak electron density that I noticed two anomalies, both of which were most strikingly evident in equatorial regions. First of all I noticed that the change-over from northern hemisphere conditions to southern hemisphere conditions occurred at the magnetic equator of the earth, and not at the geographical equator, Fig. 4. This was in striking contrast with the morphology of the lower part of the ionosphere where the ionospheric equator at, say, the equinoxes is practically coincident with the geographical equator. The second anomaly I noticed was that, during the day-time, there were two maxima of peak electron density, situated at about 20° magnetic latitude north and south of the magnetic equator, there being a minimum of the same quantity along the magnetic equator itself. Since their recognition these anomalies have been the subject of close theoretical study, and it now seems fairly certain that the two electron density maxima situated north and south of the magnetic equator are due to a large-scale 'electron-fountain' effect, the ionization falling, from a great height at the magnetic equator, downwards along the lines of magnetic force and causing concentrations at lower heights north and south of the magnetic equator. But tempting though it is for me to go into these matters in more detail, I must content myself here with stressing the fact that, even at a height of about 300 km, which is the level of the peak electron density of the ionosphere, we already have evidence of the profound influence of the earth's magnetic field in influencing the movement of ionization. When we come to consider conditions at still greater heights in the atmosphere, and even in interplanetary space beyond it, we shall encounter still further evidence of the outstanding role played by the influence of magnetic fields on the motion of electrically charged particles.

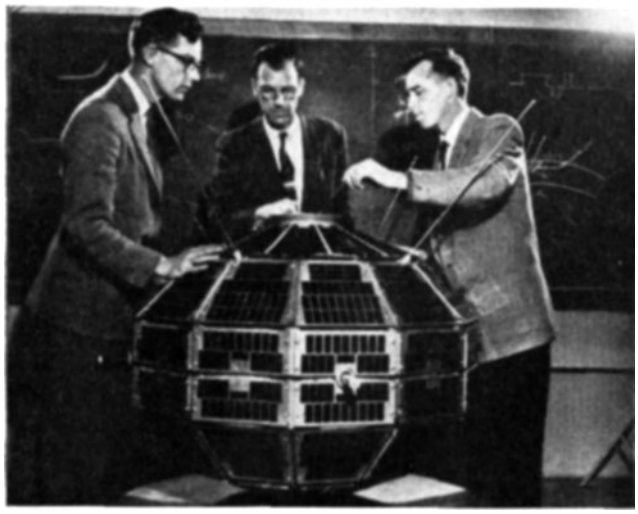


Values of F2 layer critical frequencies at noon in March 1951 plotted (*a*) as a function of geographical latitude and (*b*) as a function of geomagnetic latitude.
The two ionization maxima situated north and south of the equator illustrate what is known as the equatorial anomaly.

Fig. 4. Edinburgh University results showing that the ionospheric equator at the level of peak electron density is the magnetic equator and not the geographical equator

Now the systematic investigation of the ionosphere, using ground-based sounders operating on the radar principle, has been proceeding for over 30 years. But more recently the employment of other methods has supplemented the information available about the distribution of electron density with atmospheric height. Rocket measurements of electron density, for example, made *in situ*, have confirmed that the peak electron density in the ionosphere is reached round about the height of 300 km above the ground. But undoubtedly the most accurate and valuable information of all has been derived from the operation of what is called a 'top-side' radio sounder, situated in an earth satellite which is pursuing an orbit round the earth at a height of about 1000 km above the ground. This 'top-side' sounder consists of a complete radio pulse sender and receiver which obtains all the information necessary to make up an ionogram. But it does not record this information on a photographic film as ground-based sounders normally do. Instead it transmits the information, when prompted by the appropriate radio signal to do so, to a series of ground receiving stations situated all over the earth, Fig. 5.

Now I want to stress that the successful launching of the 'top-side' sounder, Fig. 6, which is picturesquely called *L'Alouette*, has been brought about by international co-operation. The sounder itself was conceived and constructed by Canada, and was launched by an American rocket, while the United Kingdom is assisting both Canada and the United States in receiving and interpreting the ionogram information at some of the ground stations. The three ground stations for which the United



Launched as an earth satellite in 1962 this sounder is continuing to transmit uniquely valuable information about the upper ionosphere to a network of ground stations 'on demand'.

Fig. 5. L'Alouette, *the Canadian ionospheric 'top-side' sounder (after Canadian Defence Research Telecommunications Establishment)*

Kingdom is responsible are situated at Winkfield (England), Singapore, and Port Stanley (South Atlantic).

Now the 'top-side' sounder was launched on 29th September 1962, in an approximately polar orbit: this means that it travels over the north polar region, then over the equator, then over the south polar region, and so on. Because of the low air density at 1000 km above the ground, it is expected to go on sounding the 'top-side' of the ionosphere well into next year.

It is only possible for me to give the briefest survey of the enormous wealth of information concerning the upper ionosphere which is slowly being accumulated by way of the 'top-side' sounder. I have already indicated that the terrestrial ionosphere extends to very great heights above the ground. This fact is most strikingly illustrated by the new Canadian results which show that the 'top-side' sounder, though orbiting at a height of 1000 km, is still circulating within the ionosphere itself, Fig. 7. At that level the operation of the radio sounding equipment shows that the electron population, even during the night, is still of the order of 2000 electrons per cubic centimetre. During the day-time much higher values of electron density are, of course, recorded at the same level, which reach a maximum of 40 000 electrons per cubic centimetre above the magnetic equator. In this connection it is important to note again the profound influence of the earth's magnetic field on the morphology of the 'top-side' of the ionosphere. I have already mentioned that, at the level of peak electron density, there are two anomalous maxima in the day-time ionosphere, situated 20° north and south of the magnetic equator. The results of observations made with the 'top-side' sounder have provided striking evidence in support of the 'electron-fountain' theory which, as I mentioned earlier, has been advanced to explain the formation of these maxima. It has been found, for example, that maxima of ionization still exist in the 'top-side' of the ionosphere, but their separation in magnetic latitude grows gradually less as the height above ground increases. At a height of 600–700 km the

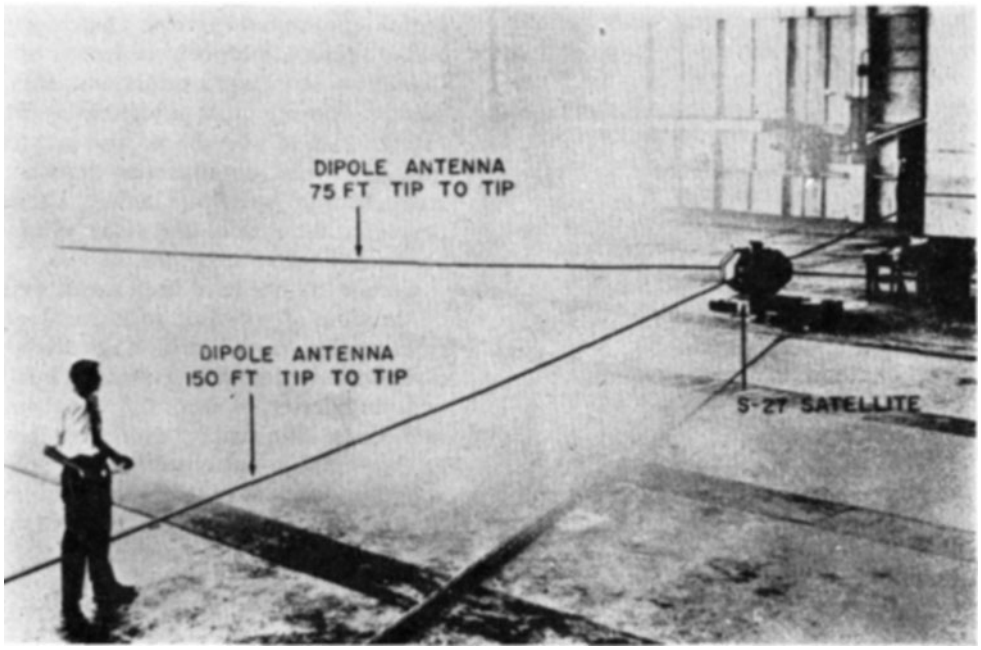


Fig. 6. The two radio aerials of the Canadian 'top-side' sounder (after Canadian Defence Research Telecommunications Establishment)

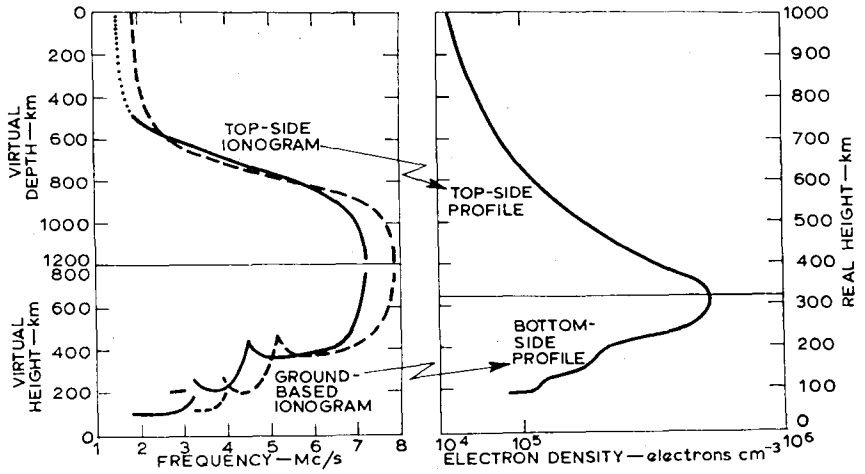


Fig. 7. Diagram showing how ionogram information from both 'top-side' and 'bottom-side' ionospheric sounders can be pieced together to derive the complete relation between electron density and height in the ionosphere (after J. O. Thomas)

bifurcation is no longer evident, there being a single maximum. But perhaps the most striking fact of all is that an imaginary line drawn through the maxima, as they occur at different heights, simulates exactly the arch-like structure of a magnetic line of force, drawn in a plane of longitude, astride the magnetic equator, Fig. 8.

I now turn to consider in greater detail the influence of the sunspot cycle on conditions in the upper atmosphere and in space beyond it. The dark markings or sunspots on the solar disc can be regarded as a species of rash, indicative of a kind of solar fever. When sunspots are large and plentiful, as they are every eleven years, the steady emissions from the sun, in the X-ray, ultra-violet and centimetric ranges of wave-length, are enhanced in intensity. The lower ionosphere, which we think is produced by solar X-rays, has an electron density—of the order of one hundred thousand electrons per cubic centimetre—which increases by 60 per cent as a period of sunspot maximum succeeds a period of sunspot minimum. The corresponding change at the level of the peak density of the ionosphere is even greater, of the order of 300 per cent. The absorption of radiations in the upper atmosphere is, however, not only responsible for the production of the ionosphere: it is also responsible for a certain amount of heating at the higher atmospheric levels. That is why the temperature of the upper atmosphere is greater at sunspot maximum than it is at sunspot minimum, Fig. 9, leading to the greater elevation of air at high levels. But sunspot regions, like the rash of a fever, can also erupt from time to time, leading to the simultaneous emission of both electromagnetic and particle radiation. These radiations are, of course, additional to the solar wind which is a steady product of the solar corona.

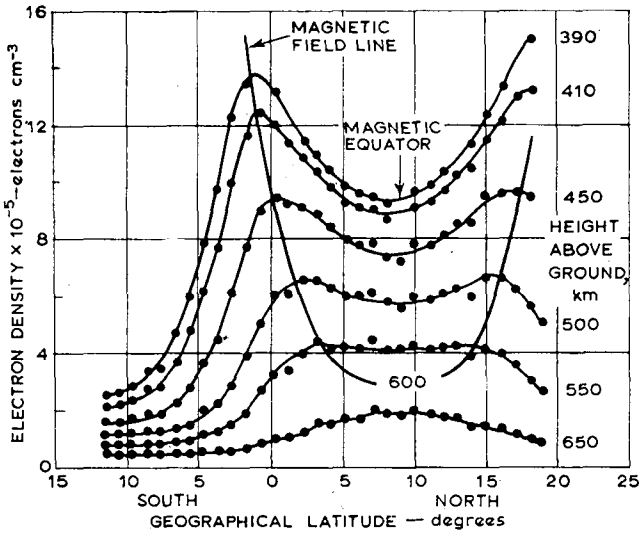


Fig. 8. The equatorial anomaly and its variation with height in the upper ionosphere; the anomaly is seen to be geomagnetically controlled (after Radio Research Station, Slough)

Careful studies have been made, over many years now, of the time of travel of solar particles over the distance from the sun to the earth. Such studies are based on the assumption that both electromagnetic waves and electrified particles erupt from the same sunspot region at the same time. Two striking events are then first noted on the earth. A visible solar flare can be seen with appropriate apparatus; and, at the same time and over the same duration, a radio fade-out is noticed over the sunlit side of the earth. We can use both events as the starting signal for the travel of material particles across space, after allowing for the eight and a half minutes required for the radiations which cause them to make the sun-to-earth journey.

Assuming such starting times for the bursts of electrified material, it has been found that they arrive at the earth in

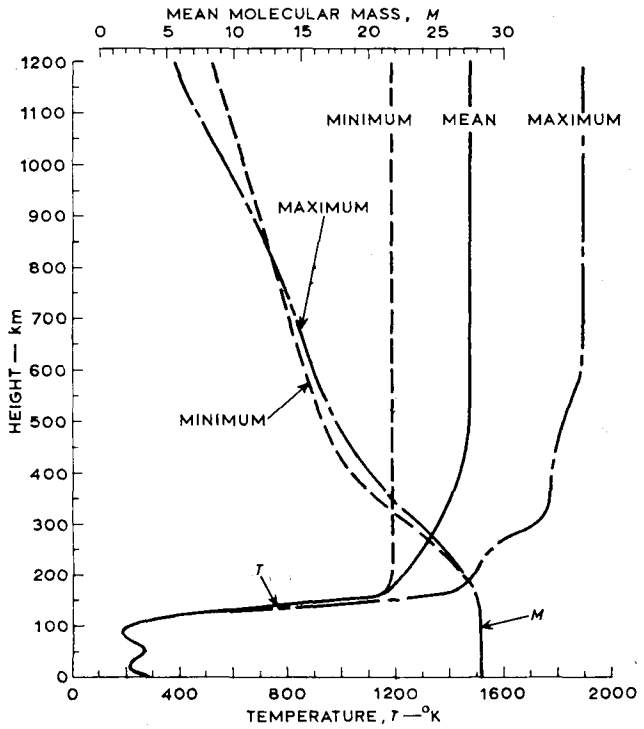


Fig. 9. Relations between atmospheric temperature and mean molecular mass with height for different degrees of sunspot activity (after Dieminger)

three main groups. There is the fastest group which perform the journey from the sun to the earth in about an hour. We call these solar cosmic rays. The second group of lower speed take about three or four hours to perform the same journey, but, because of their lower speeds, they are much influenced by the earth's magnetic field which guides them into the two polar caps of the earth. On arrival at the polar caps, where they impinge regardless of local time, they plunge into the ionosphere and produce a low-level absorbing fringe for radio waves, Fig. 10. This accounts for what is called *polar cap absorption*. The effect of such absorption in causing 'radio polar blackouts', notably during daylight hours, was first discovered by a British radio expedition during the Second International Polar Year 1933–34. But only during the last decade has its origin become clear.

The laggards among the solar particles, arriving at the earth after a solar eruption, take about thirty hours to perform the journey across space. These particles are not energetic enough to reach ground level, or even the polar caps. Instead, they fall into the atmosphere in fairly high latitudes, where they cause the familiar manifestations of auroral displays and magnetic storms.

Now I have used the word laggards to describe the group of particles which arrive late among the travelling radiations from the sun to the earth. This leads me naturally to the question of whether the late arrival of this group is attributable entirely to its low speed or whether it has been delayed, in some fashion, in the course of its journey. For we know, from one of the first discoveries of the present era of space research, that electrical particles can be trapped in the vicinity of the earth owing to the influence of the earth's magnetic field, Fig. 11. Two such naturally occurring regions of imprisoned radiation have been given the names of Van Allen belts, after their discoverer. A third belt of this kind has been deliberately, or inadvertently, created as a result of the American megaton bomb explosion at a high atmospheric level, over the Pacific Ocean, in 1962. Now there is considerable evidence to suggest that solar particles erupting from a sunspot region can find a temporary lodging in belts of this kind before they are eventually dumped into the earth's atmosphere at high latitudes. So here we may note one possible source of delay. Another would arise if there were a large-scale magnetic field in the space between the sun and the earth which deviated the particles from a rectilinear path. Such

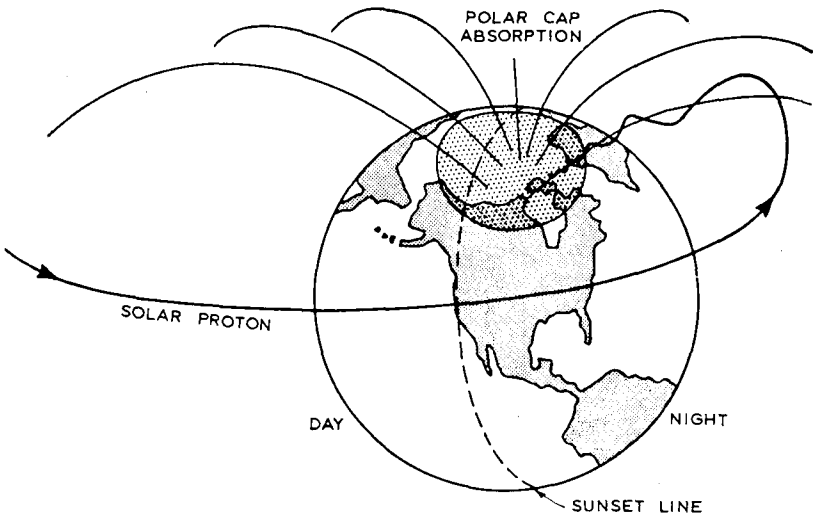


Fig. 10. Entry of solar protons into a polar region—this gives rise to polar cap absorption *(after Warwick)*

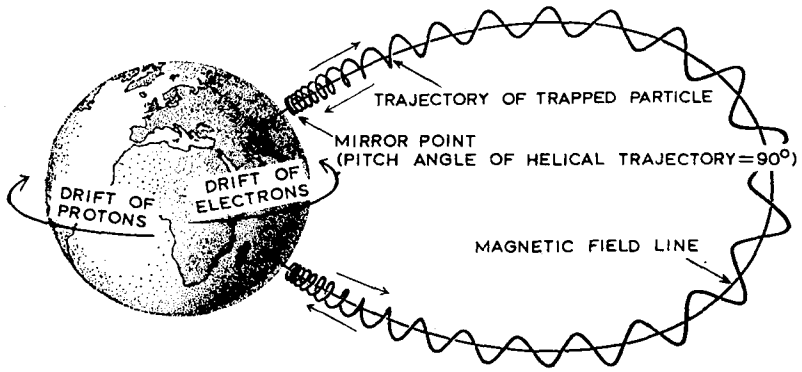


Fig. 11. How electrified particles can be imprisoned by the influence of the earth's magnetic field in a radiation belt in the vicinity of the earth (the Van Allen belts are radiation belts) (after Hess)

a magnetic field could well arise from the flow of electrified material. Theoretical work on this subject is notably difficult and no generally accepted conclusions are yet available. There are, however, two lines of experimental evidence which suggest that the space between the sun and the earth is the location of a magnetic field which is sufficiently intense to influence the travel of electrified particles. Certain species of cosmic rays are known to originate outside the solar system. The long-term recording of the terrestrial incidence of such particles has shown remarkable decreases in their intensity during magnetic storms, Fig. 12. Evidently the storm material must influence the flight of cosmic rays by magnetic deviation within the solar system. The other line of evidence has only been available since the period of the last sunspot maximum. I have mentioned earlier that fast solar protons are responsible for producing polar cap absorption. Careful measurements of the average time of travel of such protons have shown that they have completed their sun–earth journey more and more quickly as the solar activity has gradually declined since 1957–58. This seems to indicate that, as the sunspots have waned, the density of interplanetary plasma has been reduced, permitting the solar protons to make their journey to the earth in more direct fashion.

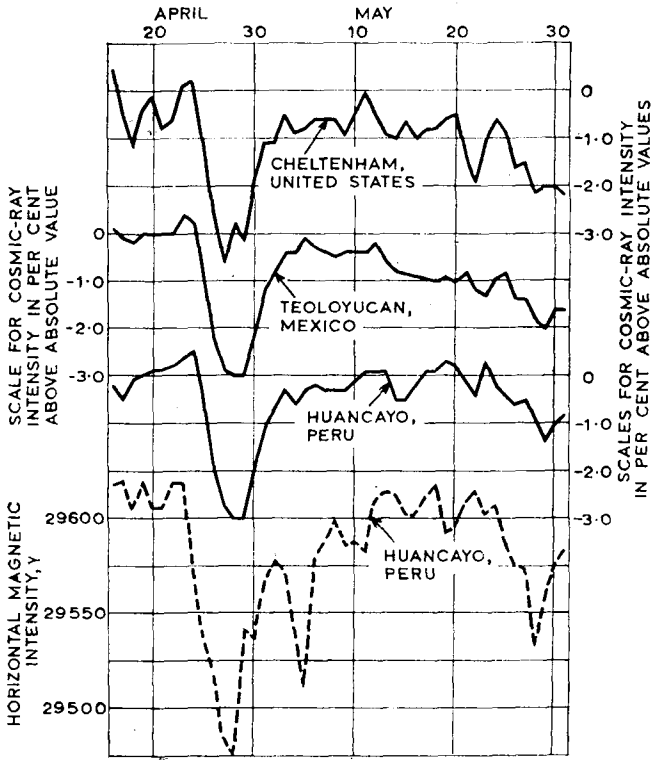


Fig. 12. Daily means of cosmic ray intensity and horizontal magnetic intensity, showing effect of magnetic storm of 25th to 30th April 1937 on cosmic-ray intensity (after Forbush)

I must now draw this brief survey of the science of near space to an end. One can, of course, be enthusiastic about the rapid advances in this field made in recent years; and I hope I shall be forgiven for stressing those aspects of that progress which specially appeal to me. But it is also abundantly clear that our knowledge and understanding of many phenomena are still incomplete; and we may appropriately recall a remark made by the late Lord Rutherford, in another connection, and say:

'This is a grand subject because there is so much in it we do not know.'